

Internet Systems Administration

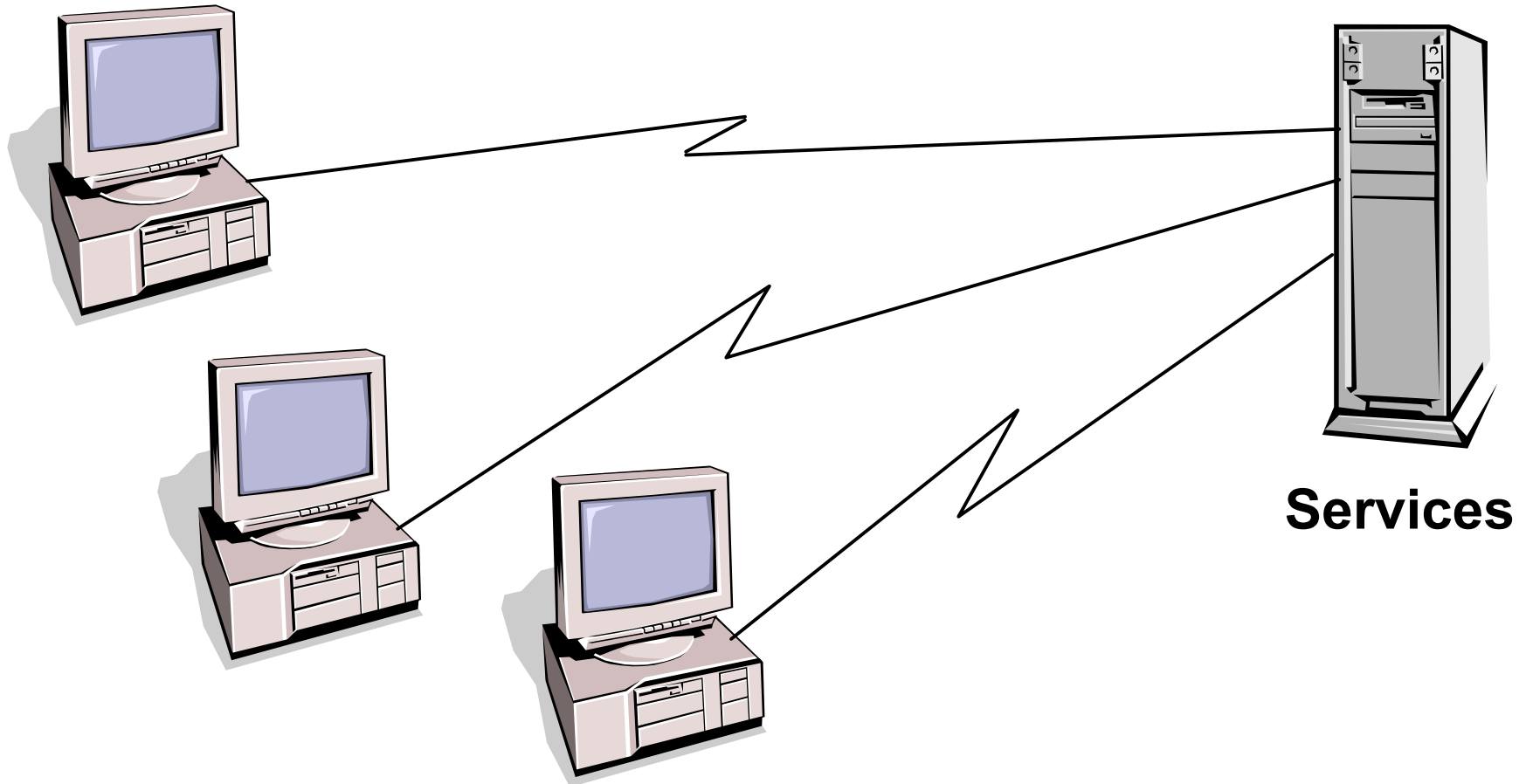
Session 1 Web Servers

Contents

- **Client/Server Basics**
- **Electronic Publishing**
- **HTTP Overview**
- **Other Web-Related Servers**

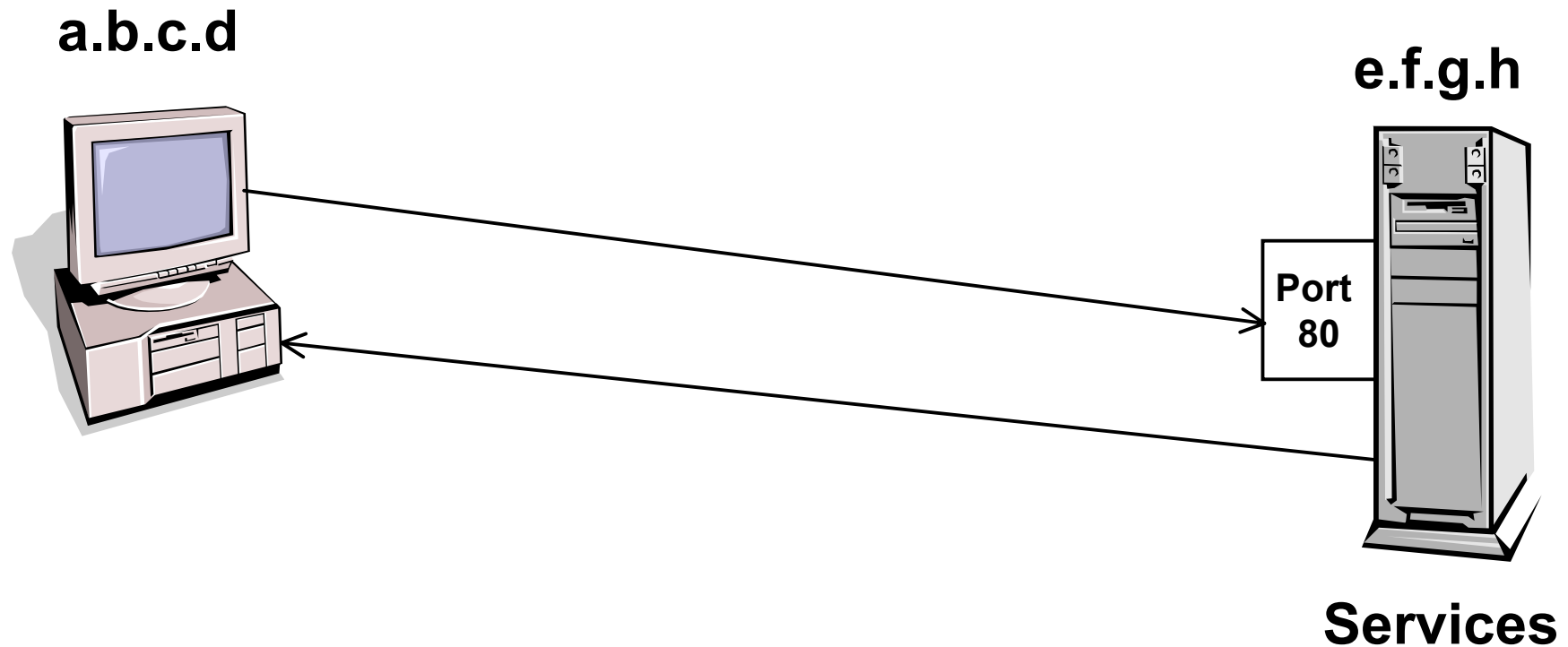
Clients and Servers

Requests

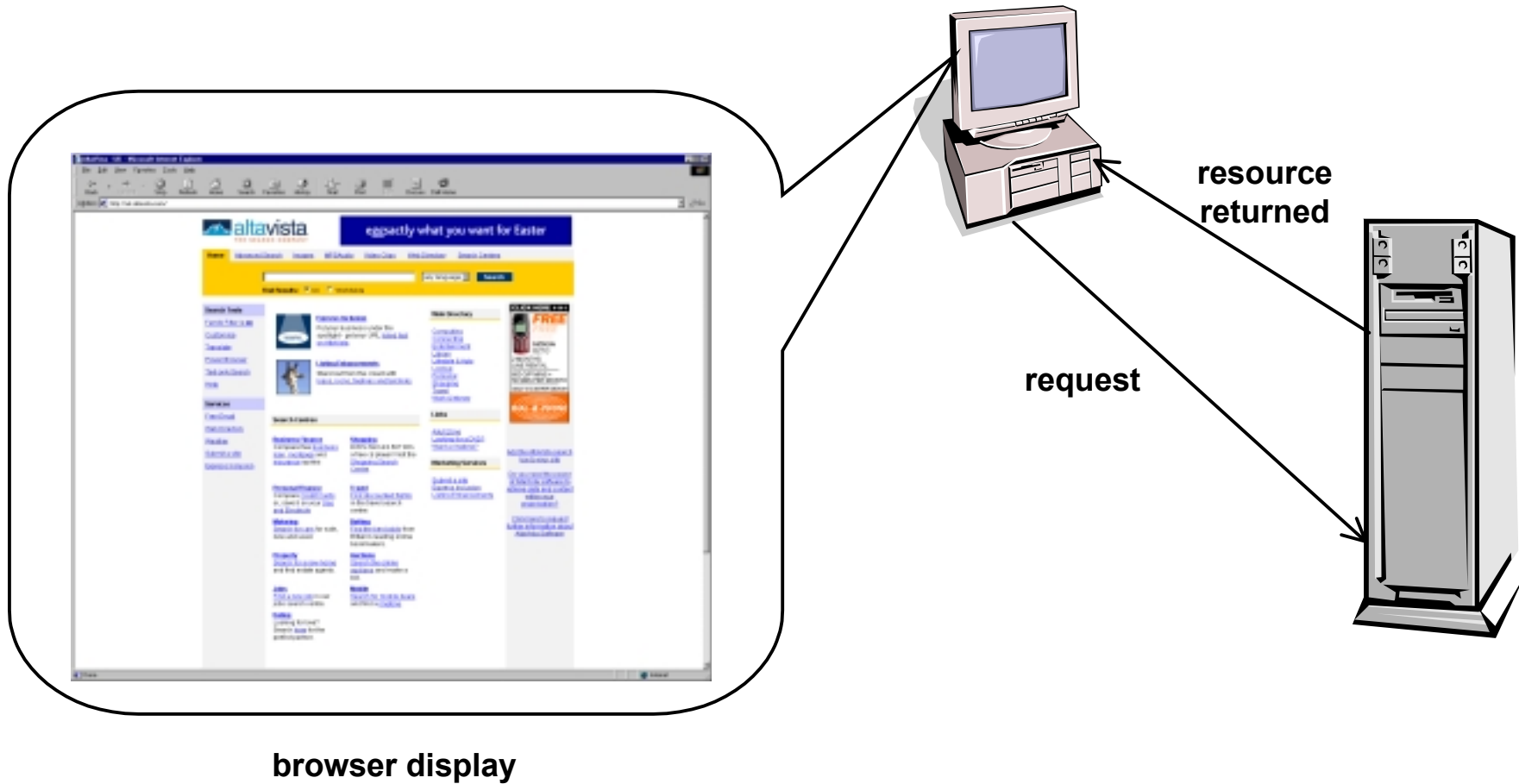


Services

TCP/IP Network Connections/Ports



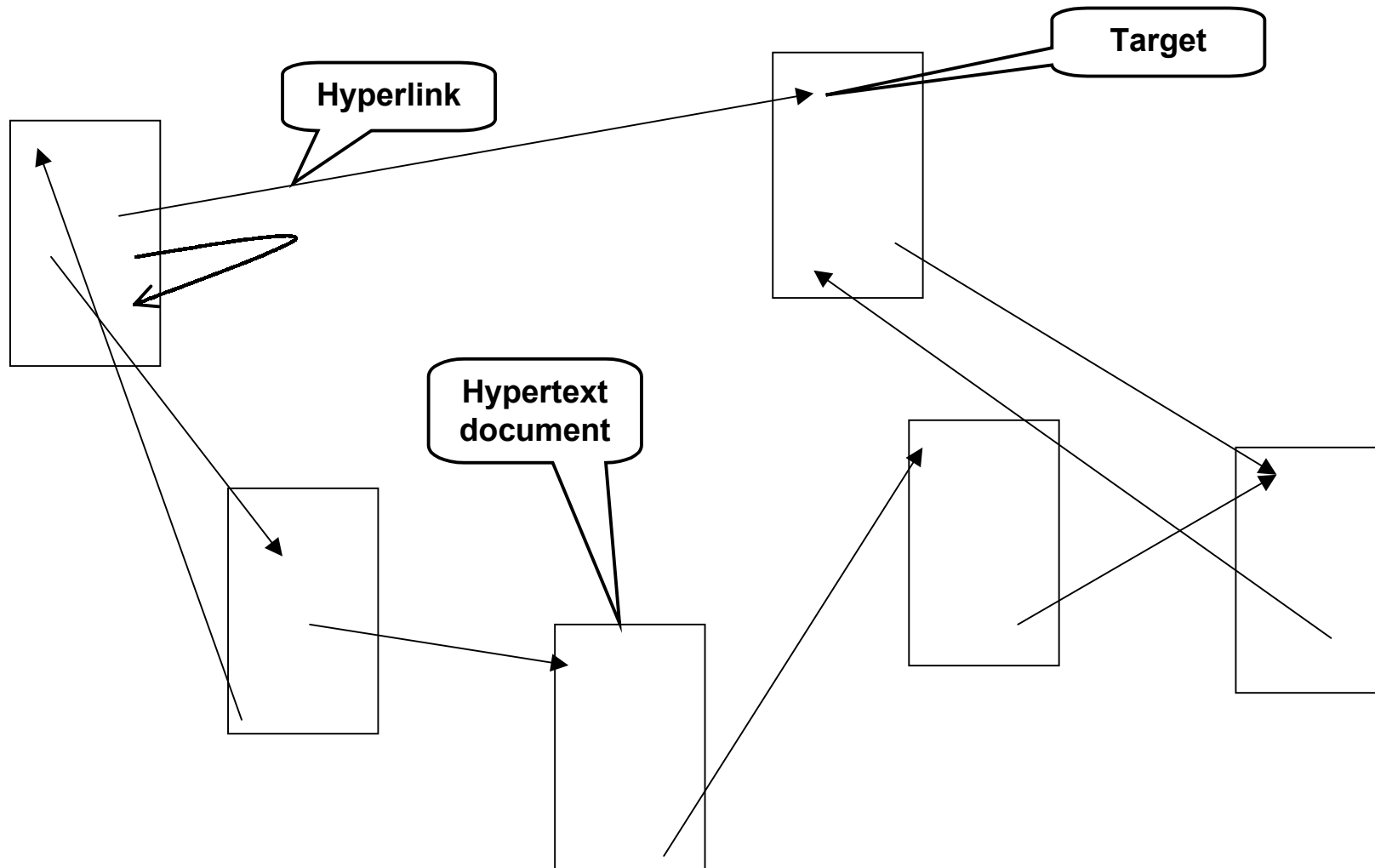
Servers and Browsers



Browsers Plug-Ins

- **Extends browser capability**
- **More than just HTML**
- **RealPlayer** **live audio and video**
- **Shockwave** **animations**
- **Acrobat Reader** **view PDF files**

Hypertext



File Types

- **ASCII text files**

- Letters number and punctuation
- View and edit with standard tools
- HTML

- **Binary files**

- Images
- Sound
- Programs

Image File Types (1)

- **GIF (Graphics Interchange Format)**
 - **256 colours**
 - **Lossless compression**
 - **Transparency**
 - **Can be animated**
 - **Good for illustrations**
 - **Proprietary (patent)**

Image File Types (2)

- **PNG (Portable Network Graphic)**
 - **As GIF, except**
 - **more colours**
 - **no animation**
 - **not proprietary**
- **JPEG (Joint Photographic Experts Group)**
 - **Millions of colours**
 - **Lossy compression**
 - **Good for photographs**

Audio File Types

- **WAV**
 - Windows
- **AIFF**
 - Macintosh
- **AU**
 - UNIX
- **Modern browsers support all these and more**

MIME Types

- **Application**
 - application/excel
- **Audio**
 - audio/midi
- **Image**
 - image/jpeg
- **Message**
 - message/news
- **Multipart**
 - multipart/digest
- **Text**
 - text/html
- **Video**
 - video/mpeg

HTTP Request

Request Line

**Request method,
resource location,
protocol version**

Header Section

**Information about
request, client**

Entity Body

**Data to be passed to the
server**

HTTP Response

Status Line

**Status code, reason
phrase, protocol version**

Header Section

**Information about server,
response**

Entity Body

**Requested resource –
often HTML**

Request Methods

● GET

- Typical way of getting a resource from a server
- Can be used to pass data to the server

● HEAD

- Server returns only header data
- Use to verify the existence of a resource

● POST

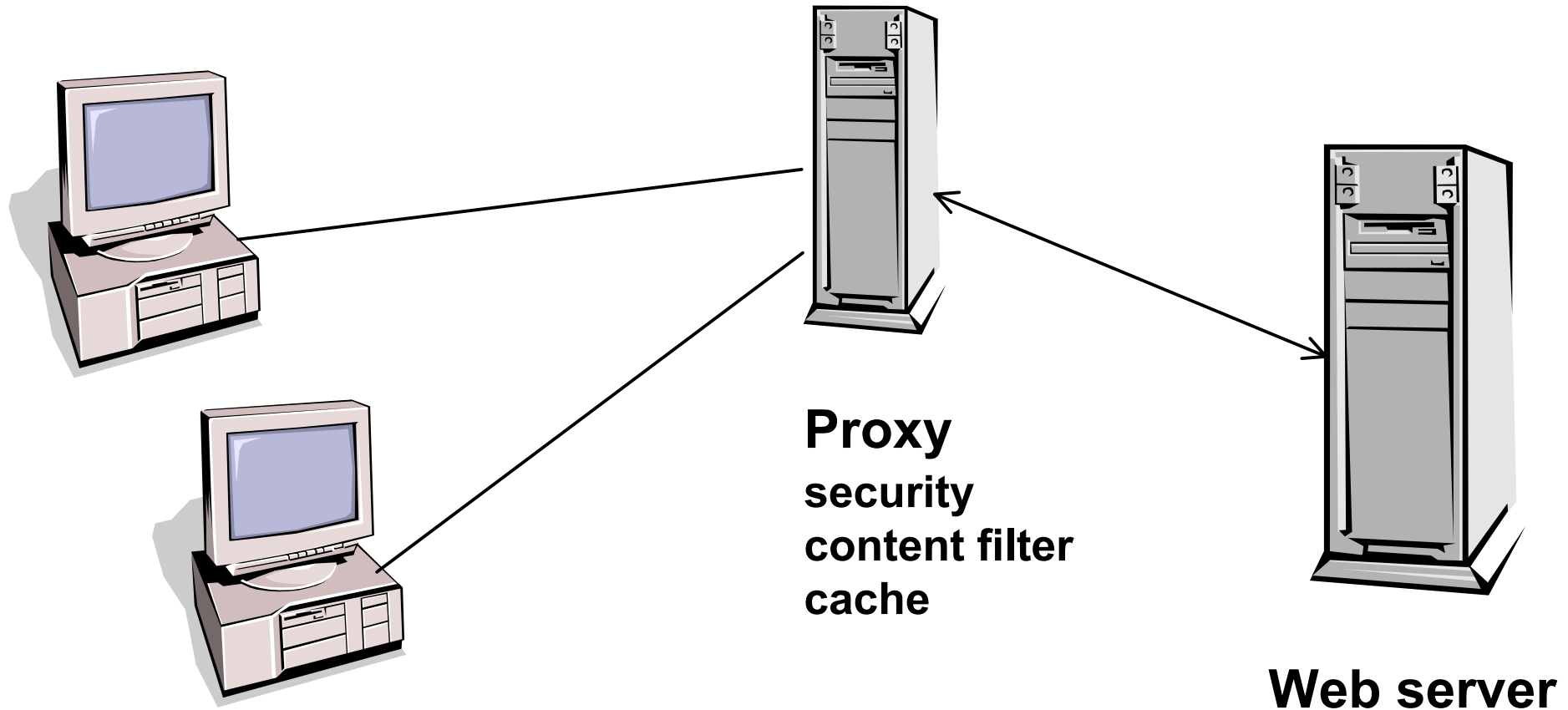
- Used to send data to the server
- Typically- send HTML form data to the server

HTTP Status Code Categories

- Informational
- Success
- Redirection
- Client error
- Server error



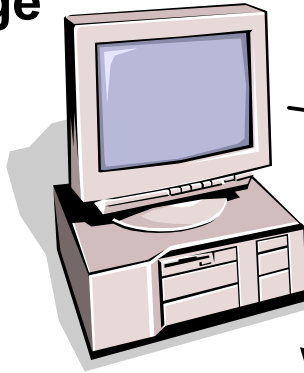
Proxy Server



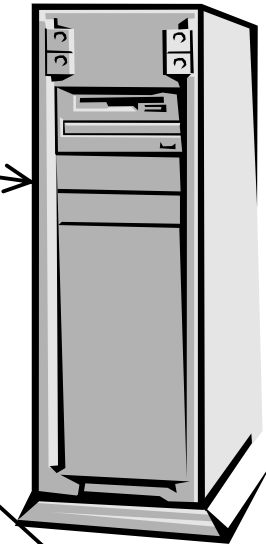
Streaming

Display without
waiting for
complete
message

3



1
Request
resource

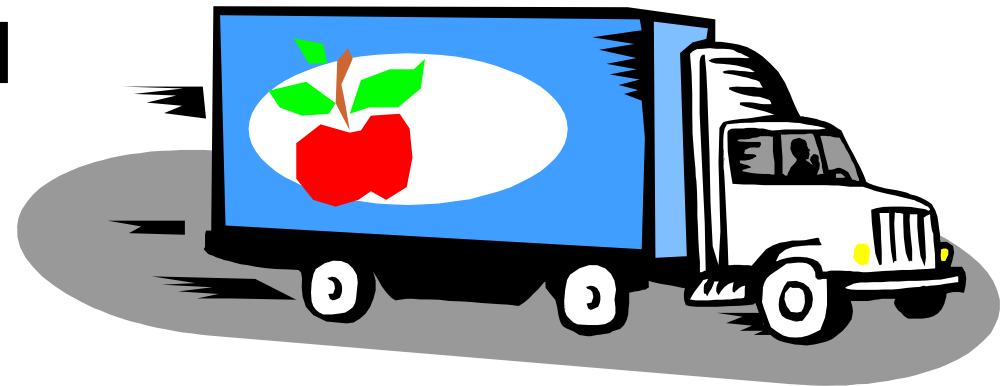


2
Data transmitted
continuously

2

FTP

- Copies files from one host to another
- Used to retrieve files from Internet archives
- Useful for binary and text files
- Log in identification
 - Anonymous ftp



SSL

- **Secure Sockets Layer**
- **Encrypts data in TCP/IP packets**
 - ordinary HTTP uses clear text
- **Commercial web applications**
- **Web server support**



Summary

- **We have covered:**
 - **Client/Server Basics**
 - **Electronic Publishing**
 - **HTTP Overview**
 - **Other Web-Related Servers**

Internet Systems Administration

Session 2

Planning a Server

Contents

- **Hosting Options**
- **UNIX or NT?**
- **Sizing a Server**
- **Domain Names**

Hosting

- **Host – a computer connected to the Internet with an address**
- **A web server is a host**
 - **Where can you locate your web server?**
 - **How can you connect your web server to the Internet?**

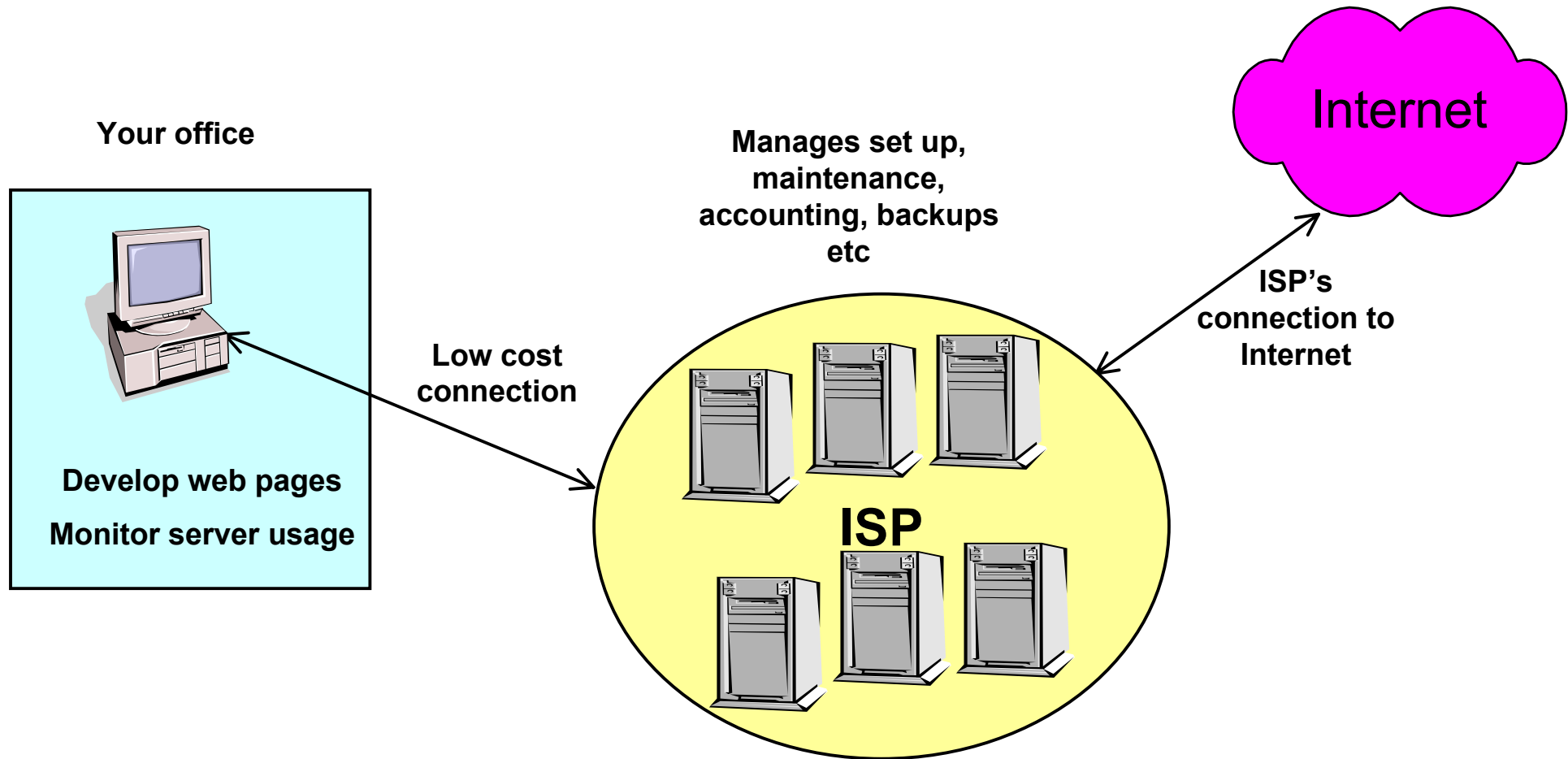
Hosting Options

- **Set up your own web server**
- **Co-location**
- **Virtual host**
- **Personal-Page sites (ISP)**
- **Free-Page sites**

Your Own Server - Issues

- **Cost of server hardware and software**
- **Operations –**
 - Backup
 - 24/7
 - Power supplies
- **Security**
 - Protecting your server
 - Protecting other peoples' resources

Co-Located/Dedicated Server



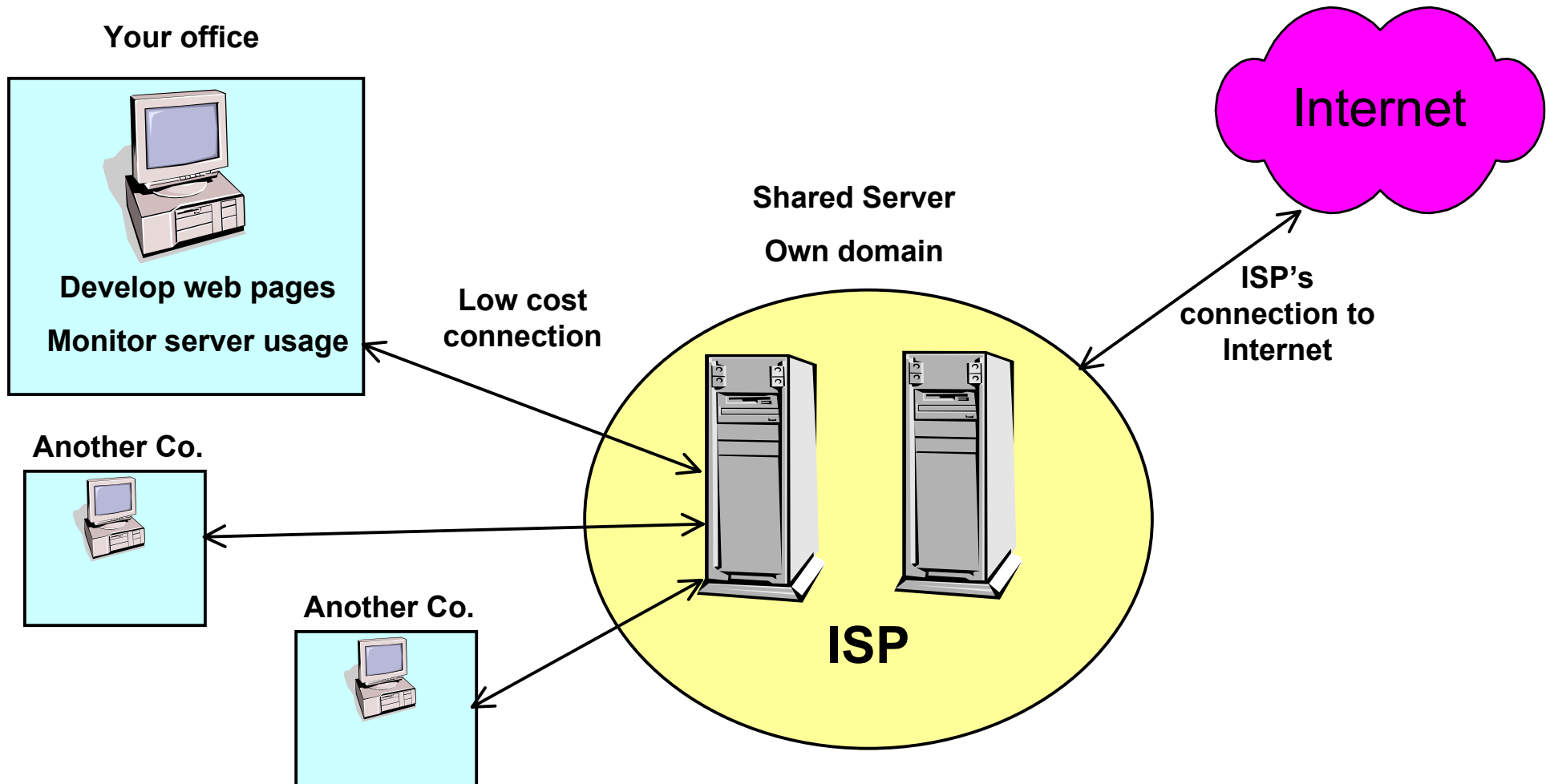
Co-Located Server Issues

- You must supply the server
- You must administer
- Fees to ISP for service
- But
 - Good connectivity to Internet
 - No local floor-space required

Dedicated Server Issues

- **Not your own server**
- **Fees to ISP for service**
- **But**
 - **Much easier to set up**
 - **24/7 support**
 - **Good connectivity to Internet**

Virtual Host



Virtual Host Issues

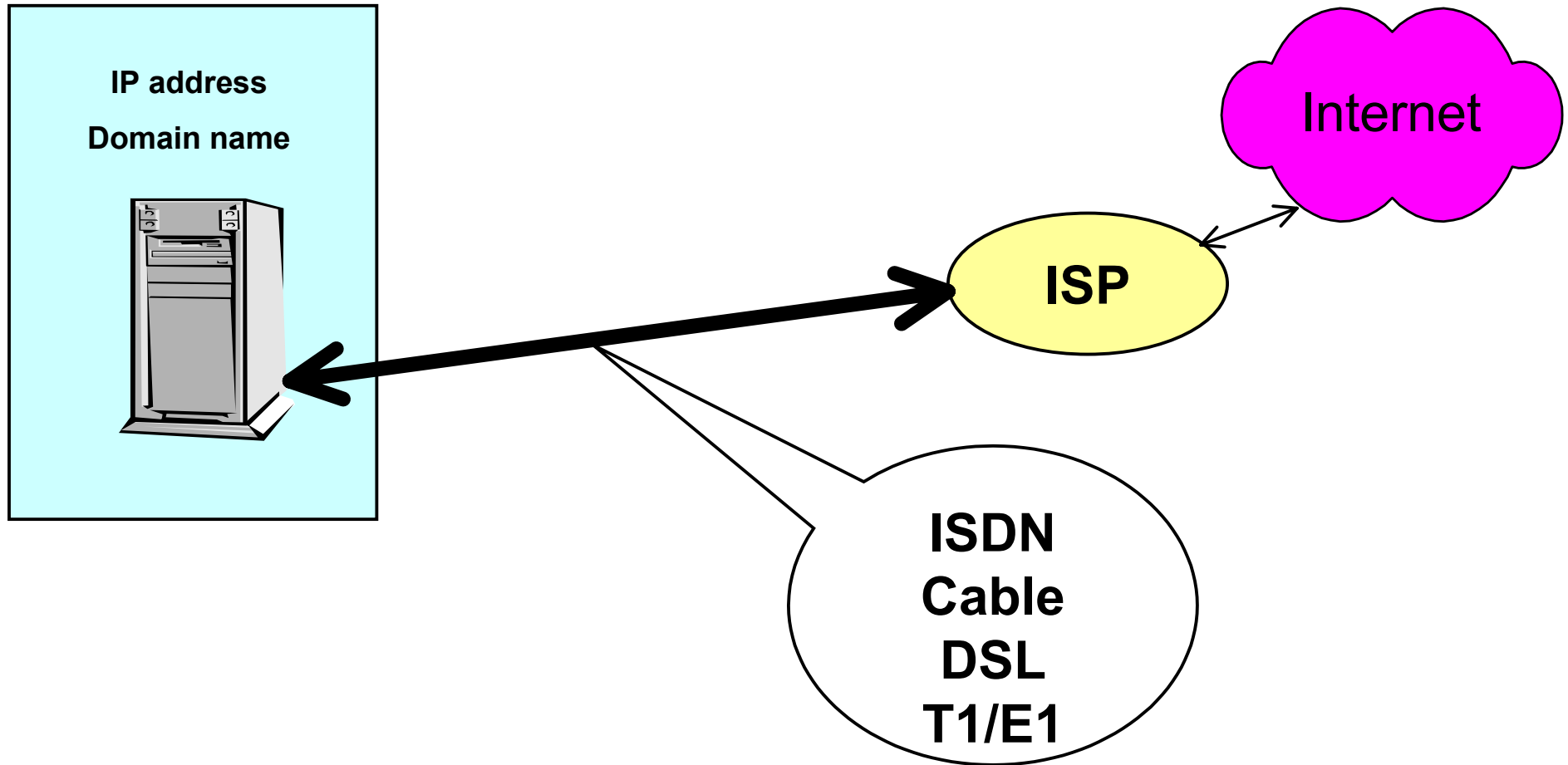
- **Shared server**
- **Limited or no server programming access**
- **Standard solutions**
- **But**
 - **Inexpensive**
 - **Good connectivity to Internet**

Personal and Free Pages

- **ISP standard offering**
- **Very limited storage**
- **Not for commercial use**
- **May be free with some ISPs**

Connecting Your Server

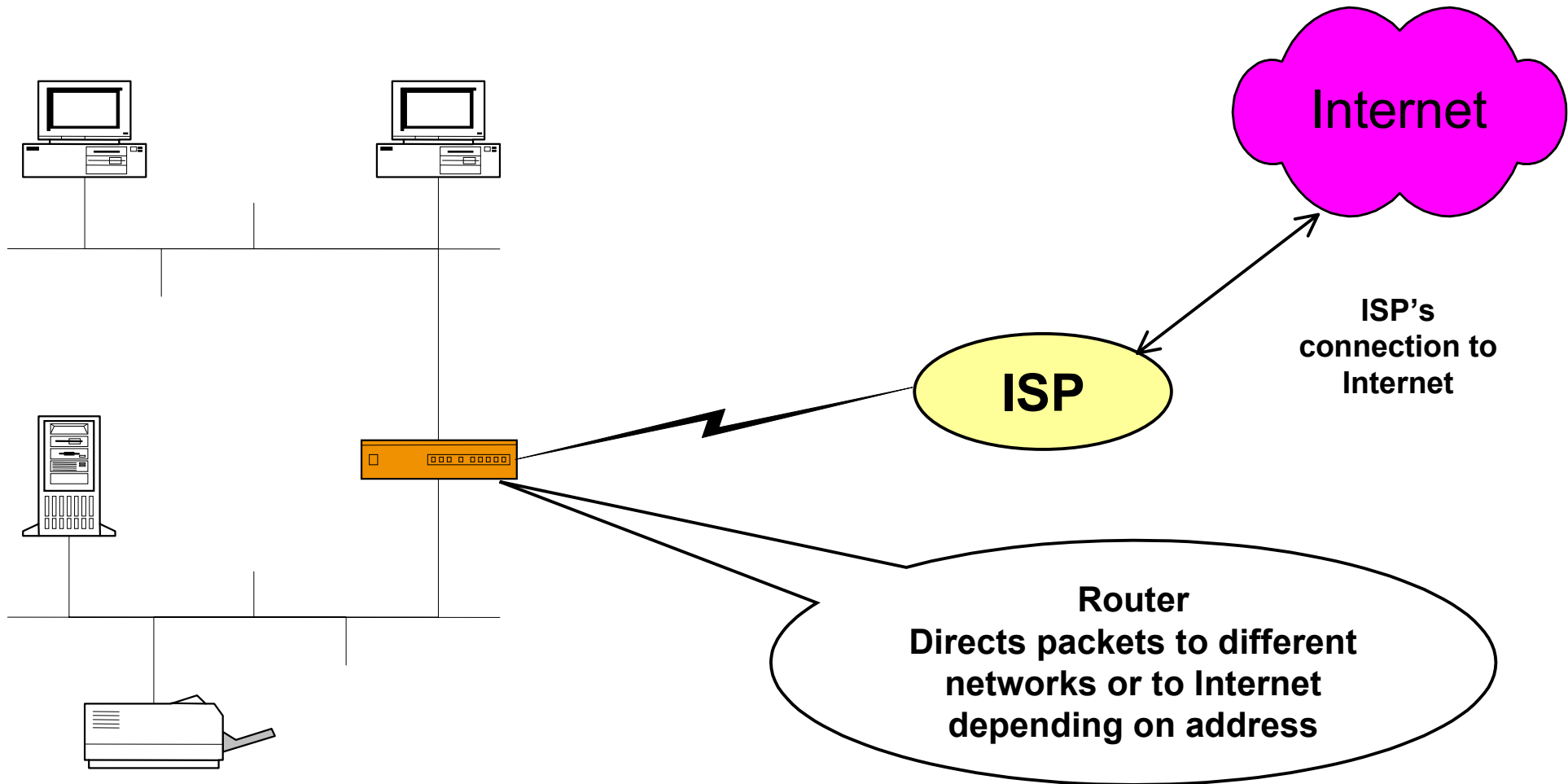
Your office



IP Addressing

- **Static IP address**
 - needed for a web server
 - never changes
 - allows other domains to connect to your server
- **Dynamic IP address**
 - different on each dial-up
 - acceptable for dial-up Internet access
 - inappropriate for running web based services

Router



Servers - UNIX or NT? (1)

● Unix

- Built for TCP/IP networking**
- Scaleable**
- Many hardware platforms**
- Robust**
- Command Language Interface**
- X Windows GUI**
- Requires longer learning curve**

Servers - UNIX or NT? (2)

● NT/XP/Win 2000

- Closed - proprietary**
- Limited hardware platform choice**
- GUI oriented**
- Easy to learn**
- Getting more robust**

Servers- UNIX or NT? (3)

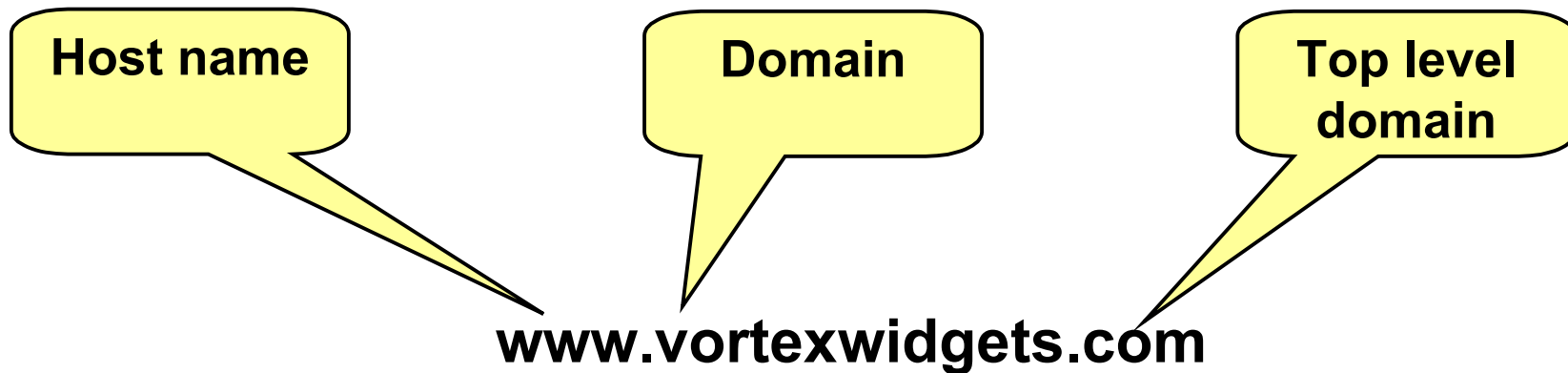
●Linux

- Open – Free version of UNIX**
- Works well on limited hardware**
- Robust**
- Very versatile**
- Requires investment in learning**
- Becoming extremely popular for web servers**

Sizing Your Server

- **Bandwidth and network capacity**
 - Moving data from and to your server
 - Buy a bigger pipe
- **Web server processor**
 - Serving out web pages fast enough to keep up with demand
 - Buy a faster CPU
 - Add RAM

Domain Names



Top Level Domains

- | | | | |
|--------|------------|-------|----------------|
| ● .com | business | ● .us | United States |
| ● .org | non-profit | ● .au | Australia |
| ● .net | ISPs | ● .uk | United Kingdom |
| ● .edu | university | ● .jp | Japan |
| ● .gov | government | ● .sw | Sweden |
| ● .mil | military | | |

Summary

- **We have covered:**

- **Hosting Options**
- **UNIX or NT?**
- **Sizing a Server**
- **Domain Names**

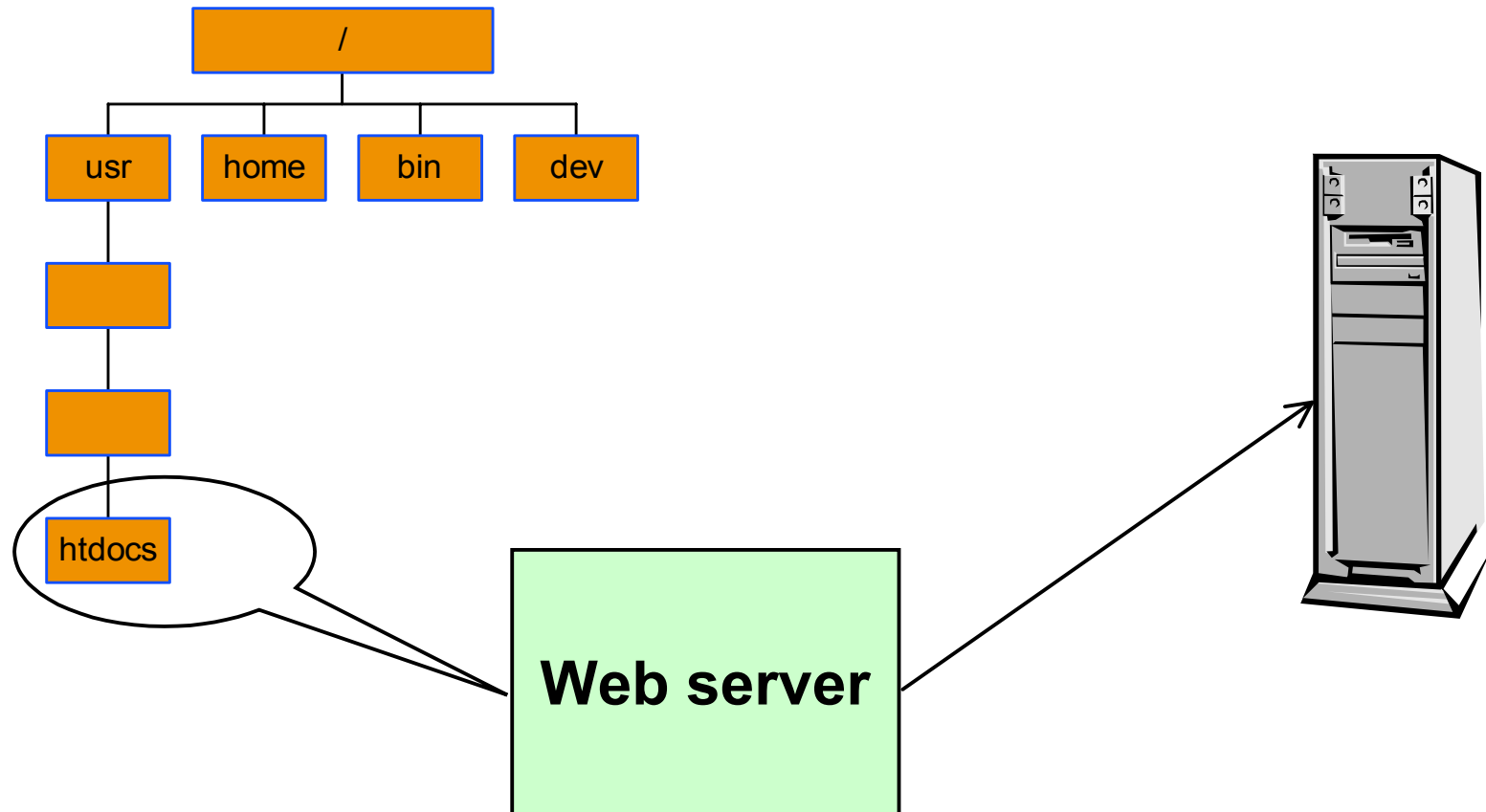
Internet Systems Administration

Session 3 Users and Documents

Contents

- **Server Users and Directories**
- **Server Administrators**
- **Document Hierarchy**
- **Directory Indexing**
- **File and Directory Names**
- **Transferring Files**

Web Servers and Directories



Users and Accounts

- **Web site visitors do not need accounts**
- **User accounts define server privileges**
- **Users have a home directory**
- **User name and password**

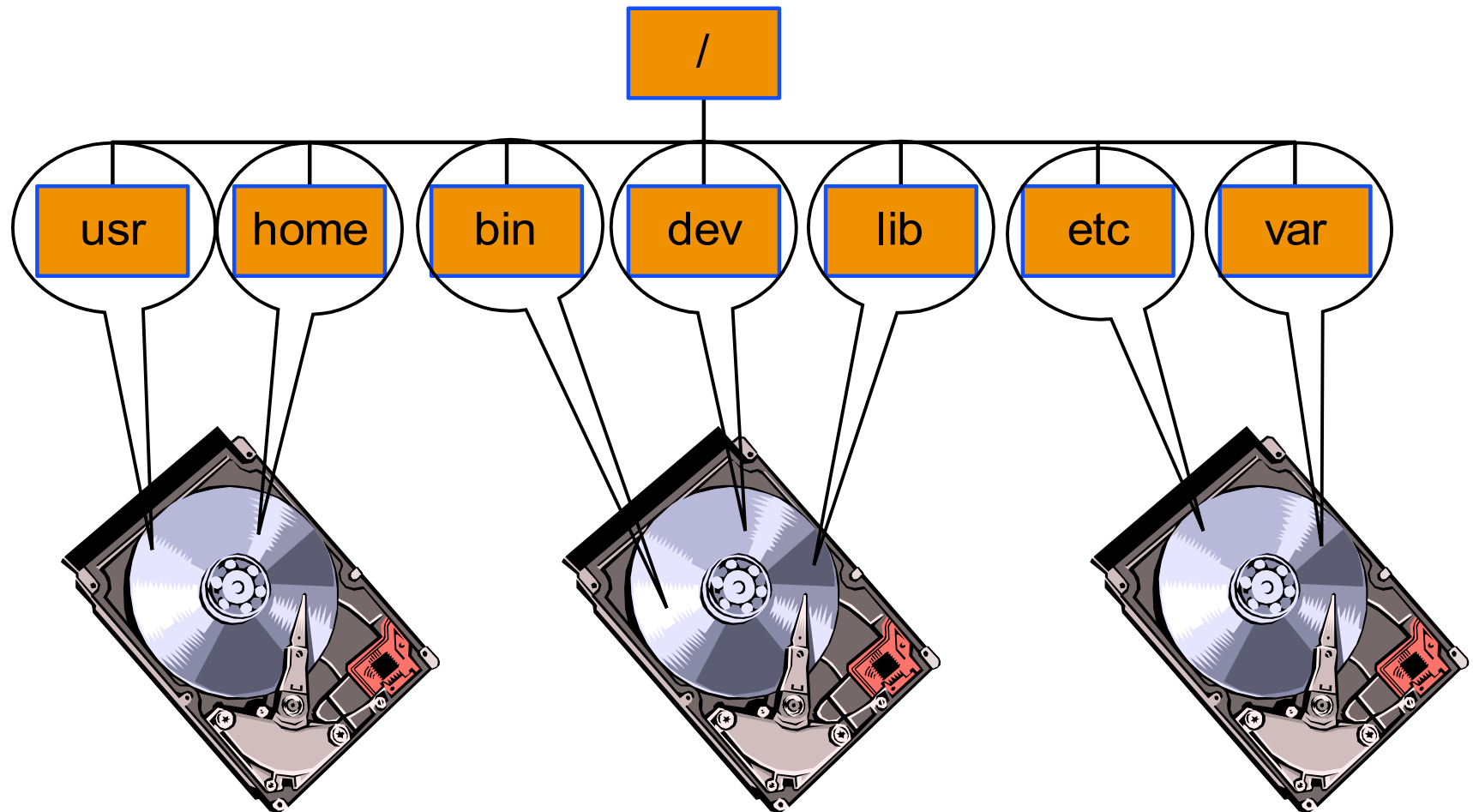
System Administration Duties (1)

- **Install and configure server OS**
- **Set up daemons/services (UNIX/NT)**
- **Install and configure web server**
- **Keep server software up to date**
 - **Patches (UNIX)**
 - **Service packs (NT)**

System Administration Duties (2)

- **Backup and recovery**
 - To disk or tape
 - Full and incremental backups
- **Accounts and quotas**
 - Maintain users and their accounts
 - Decide level of resources for users
- **Network software configuration**
- **Monitoring security**

File Systems (1)



File Systems (2)

- **Microsoft**

- **FAT/FAT16/FAT32**
- **NTFS**

- **Apple**

- **HFS**

- **UNIX**

- **UFS**
- **EXT2**
- **Reiserfs**

- **NFS**

Directories

- **Absolute Pathname**
`/usr/local/httpd/htdocs`
- **Relative pathname**
`cd /usr/local/httpd`
`cd htdocs`
- **Dot and Dot Dot**
 - `.` = current directory
 - `..` = parent directory

Uniform Resource Locators

- Describe how to find a web resource

<http://www.vortexwidgets.com/support/industrial.html>

Servername: www.vortexwidgets.com

Path: [support](#)

Filename: [industrial.html](#)

- Use relative URLs in your own web pages

- Make it easy to move to another directory

- Dot and Dot Dot

- . = current directory

- .. = parent directory

Directory Indexing

- **If no file name in a URL:**
 - **Directory browsing enabled**
 - **If index document, return default document**
 - **Otherwise return list of files**
 - **Directory browsing disabled**
 - **If index document, return default document**
 - **Otherwise return nothing**

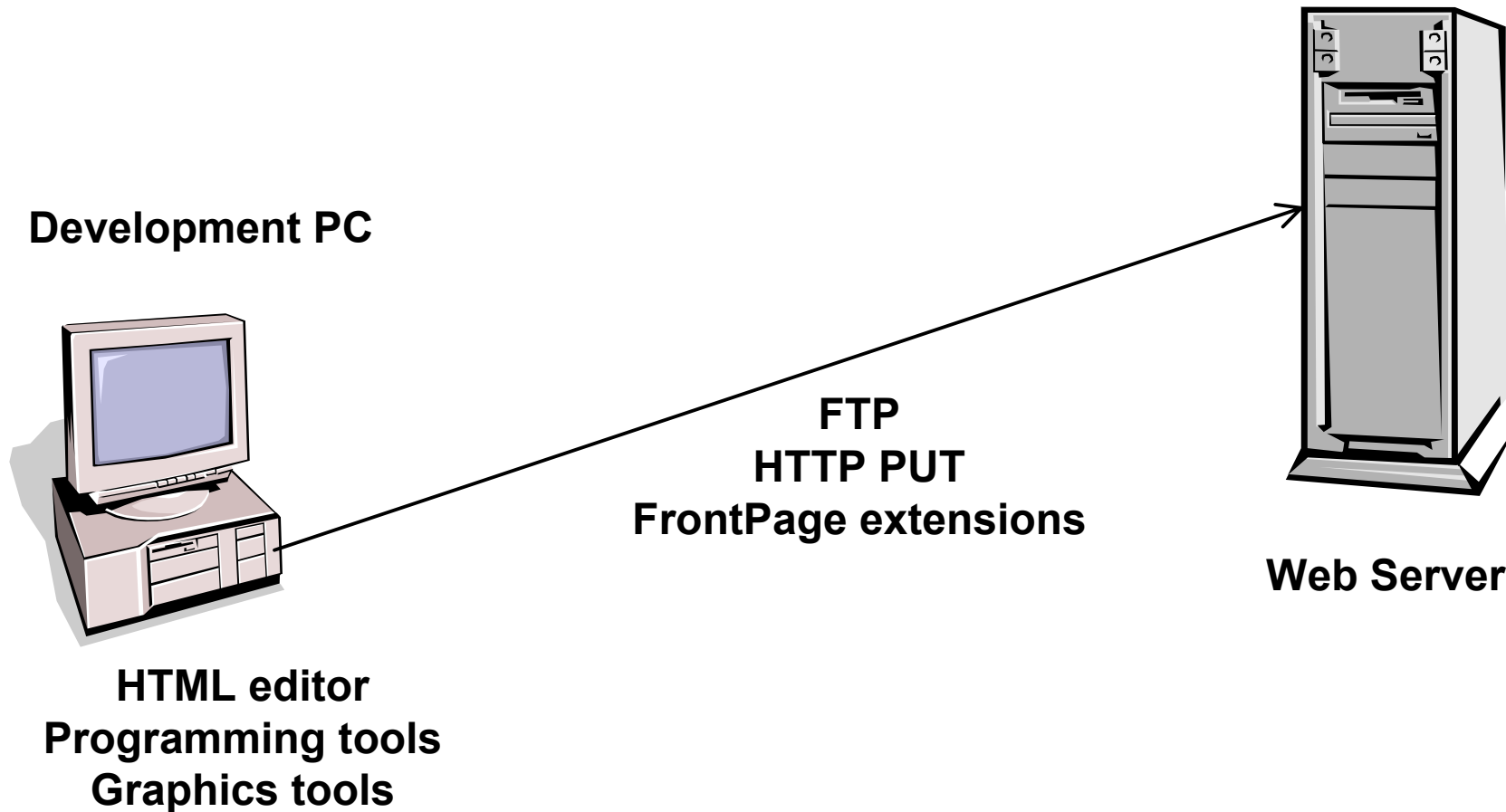
Good Filename Practice (1)

- **Don't use spaces in names**
 - Use underscores (_) or dash (-) instead
- **Don't use special characters**
& + ?
- **Keep filenames short**
- **Use a standard naming convention**

Good Filename Practice (2)

- **Use consistent filename extensions**
 - .html**
 - .gif**
- **Don't use extensions with directory names**
- **Use lowercase letters in all filenames**
 - **UNIX is case sensitive**
 - **Windows is (mostly) not case sensitive**

Transferring Files



FTP

- **File Transfer Protocol**
- **Use a GUI FTP client, OR**
- **FTP CLI Commands**

Open *hostname*

Put *filename*

Get *filename*

CD *directory*

LS (list directory)

MKDIR *dirname*

MPUT

MGET

FrontPage

- **An easy to use HTML editor**
- **Synchronisation between development PC and web server**
- **BUT**
 - **Not all ISPs support this type of usage**
 - **Works best with Microsoft web server**
 - **Proprietary**

Summary

- **We have covered:**
 - **Server Users and Directories**
 - **Server Administrators**
 - **Document Hierarchy**
 - **Directory Indexing**
 - **File and Directory Names**
 - **Transferring Files**

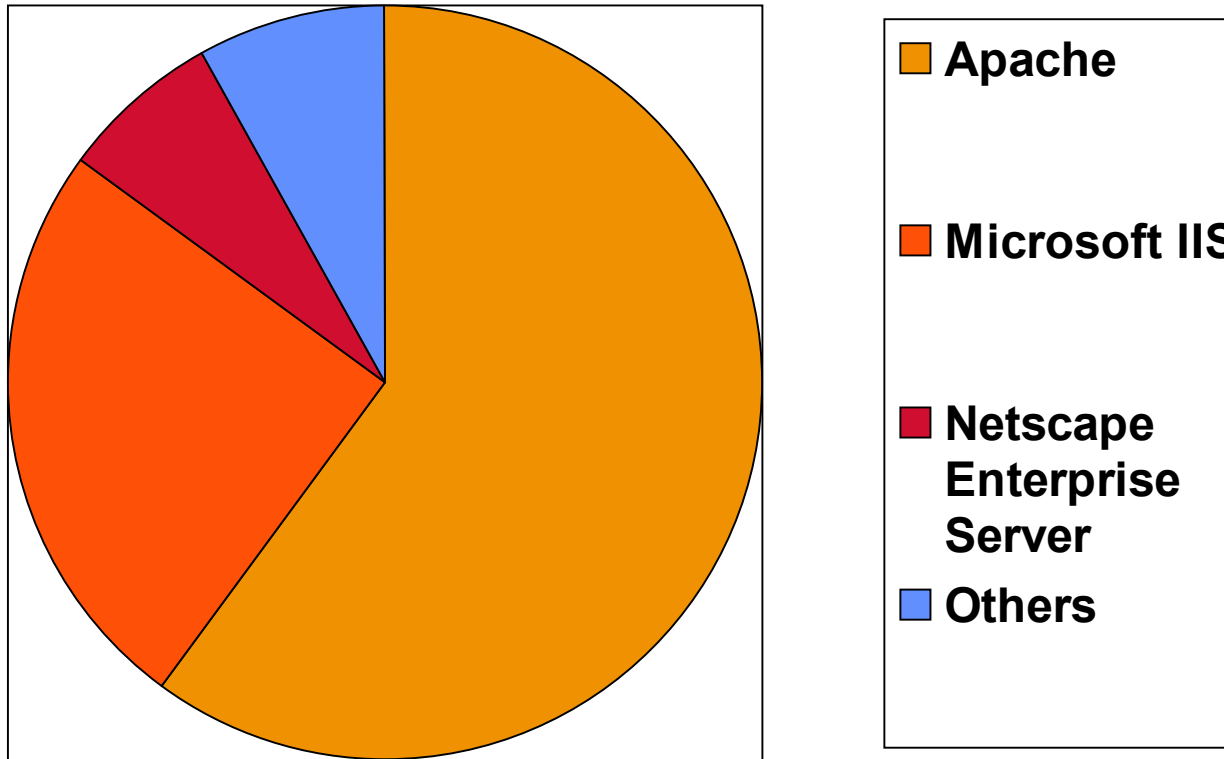
Internet Systems Administration

Session 4 Server Configuration

Contents

- **Choosing web server software**
- **Customising your web server**
- **Controlling access**
- **Secure Socket Layer configuration**
- **Virtual hosts**

Popular Web Servers



Apache

- **Open Source**
- **Multiple platforms (UNIX and Microsoft)**
- **Very powerful and configurable**
- **Uses configuration files**
`httpd.conf`



Microsoft IIS

- **Easy to use, GUI oriented**
- **Closed - proprietary**
- **Microsoft Management Console**
- **Extendable through ISAPI**
 - **DLL**
 - **ASP**
- **Support for FrontPage extensions**

Evaluating Web Servers

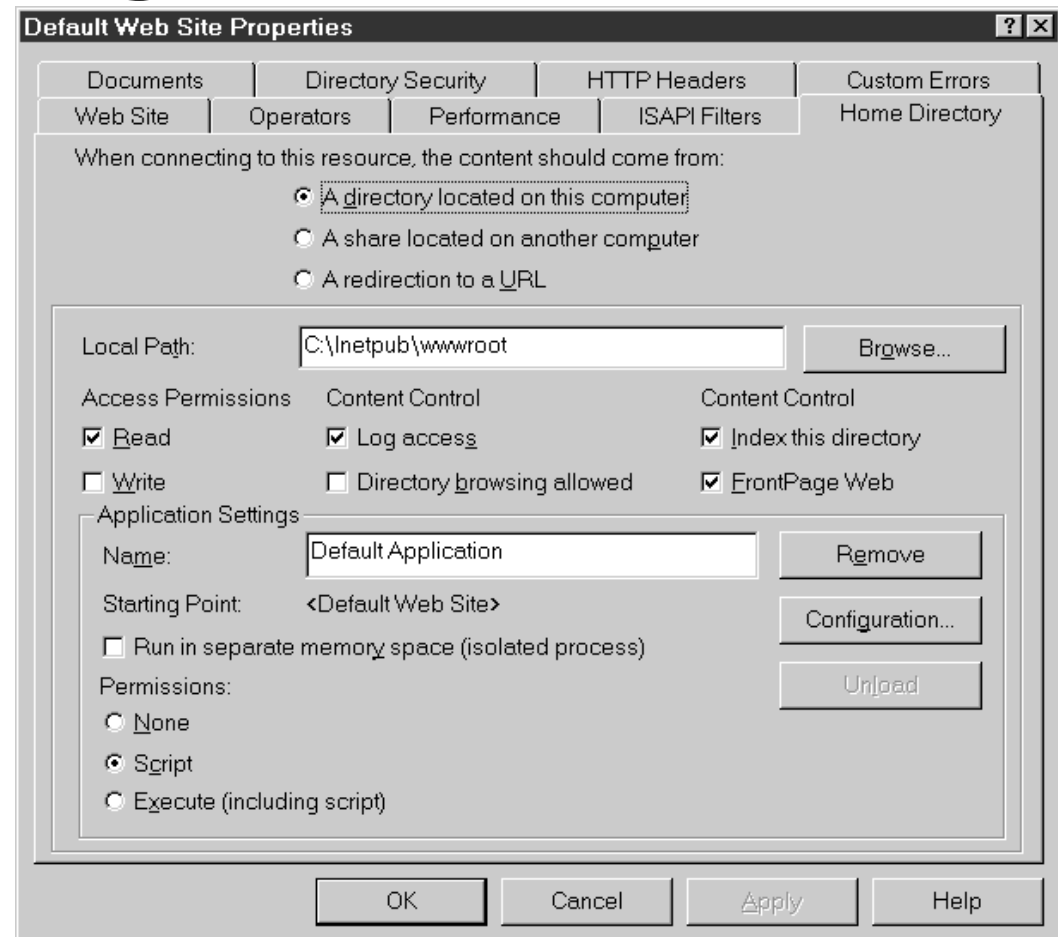
- **Performance Benchmarks**
 - SPECweb96/99
 - WebStone
- **Cost**
- **Ease of use or installation**
- **Scalability**
- **Security**
- **Support of industry standards**

Apache Configuration

- **All configuration through configuration files**
- **Directives define options**
- **Directives are organised into sections:**
 - Directory
 - DirectoryMatch
 - Files
 - FilesMatch
 - Location
 - LocationMatch

IIS Configuration

- IP address
- TCP port
- Home directory
- Execute
- Virtual directory
- Default document
- Directory browsing
- Authentication control
- Application mappings
- Redirect to URL



Controlling Access

- **UNIX model**
- **File permissions**
 - User, Group, Other
 - Read, Write, Execute
- **In any combination per file**
 - and directory

HTTP Authentication

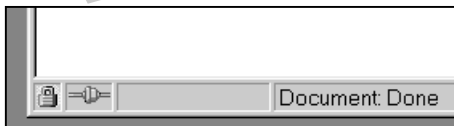
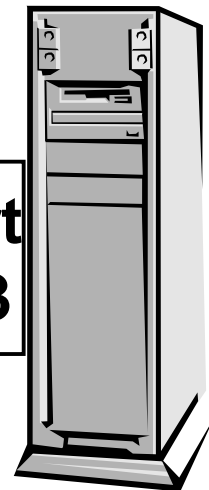
- **Protect specific files and directories**
- **Require User Name and Password**
- **HTTP1.1**
 - Basic authentication
 - no encryption across network
 - Digest authentication
 - uses MD5 encryption

Secure Socket Layer

HTTPS
HTTP using SSL

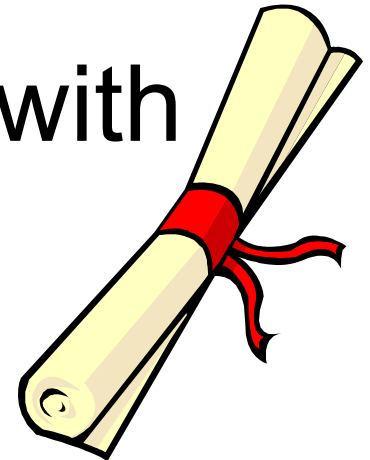
**Port
443**

Server

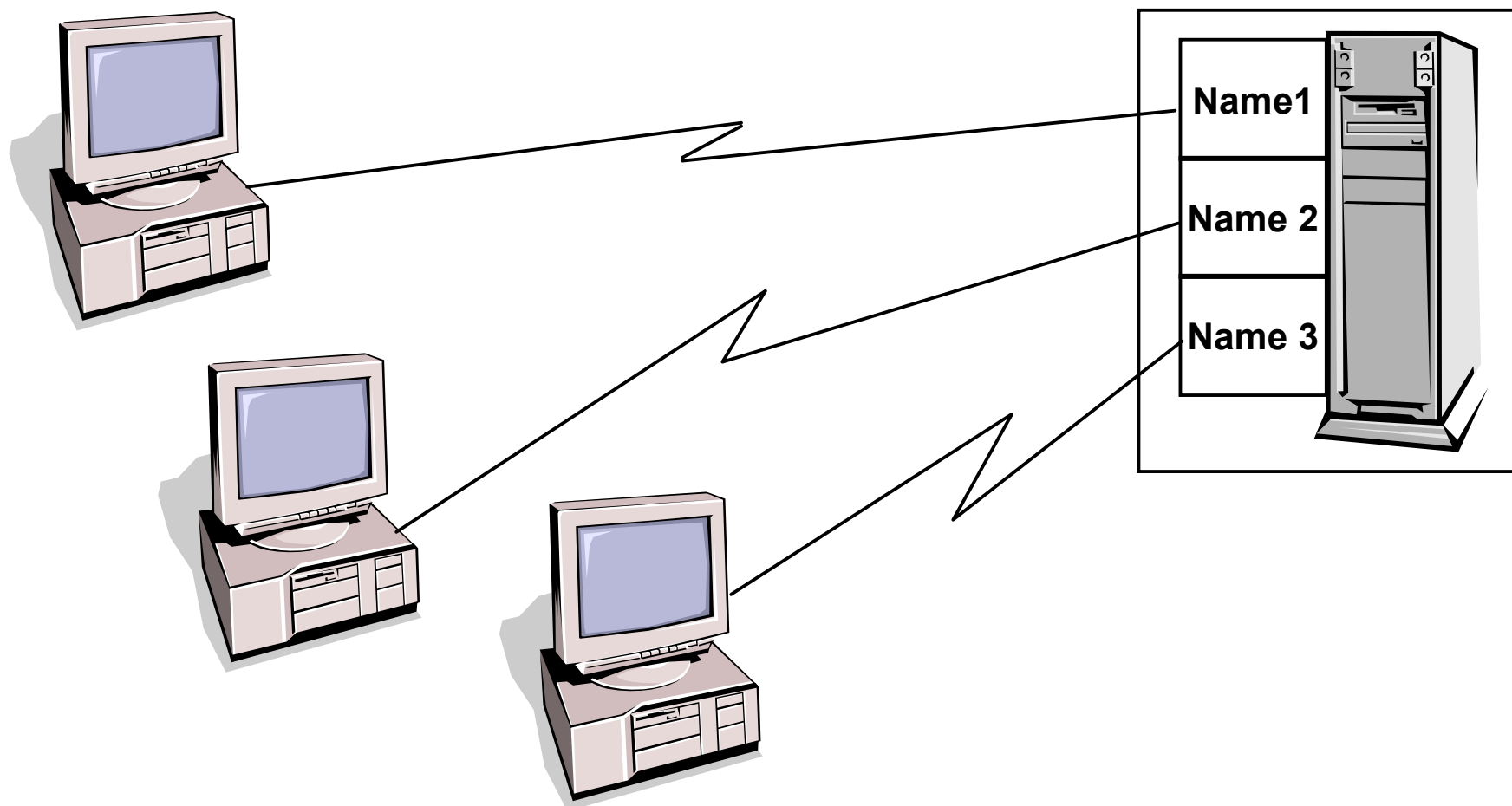


Certificates

- Supplies information about your site
- Certificate Authority issues
 - Verisign
 - Thawte
- Helps user trust a transaction with your server
- Costs?



Virtual Hosts (1)



Virtual Hosts (2)

- **Name based**

- One IP address, multiple names
- Web server distinguishes using hostname

- **IP address based**

- Multiple IP addresses (and names)
- Not all servers support this

Summary

- **We have covered:**
 - **Choosing web server software**
 - **Customising your web server**
 - **Controlling access**
 - **Secure Socket Layer configuration**
 - **Virtual hosts**

Internet Systems Administration

Session 5

Server-Side Programming

Contents

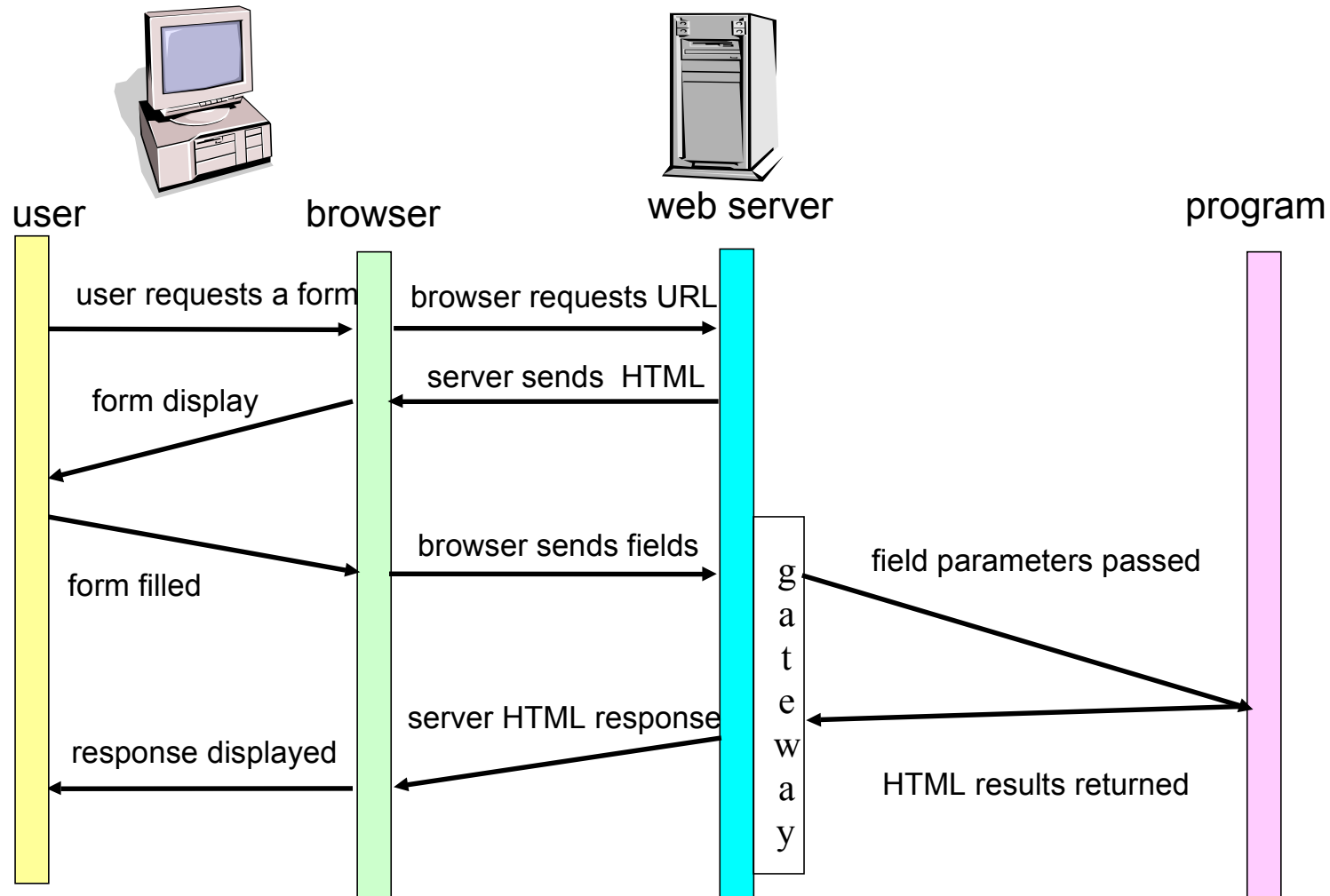
- **Dynamic Documents**
- **CGI and Forms**
- **Server-Side Includes**
- **Active Server Pages**
- **Servlets and Java Server Pages**

WHY Server-Side Programming?

- **Normal HTML is static**
 - only changes when you edit it
- **Some web sites need to present information that changes rapidly**
- **Some web sites need to process data received from visitors**
- **Server-Side Programming meets this requirement**



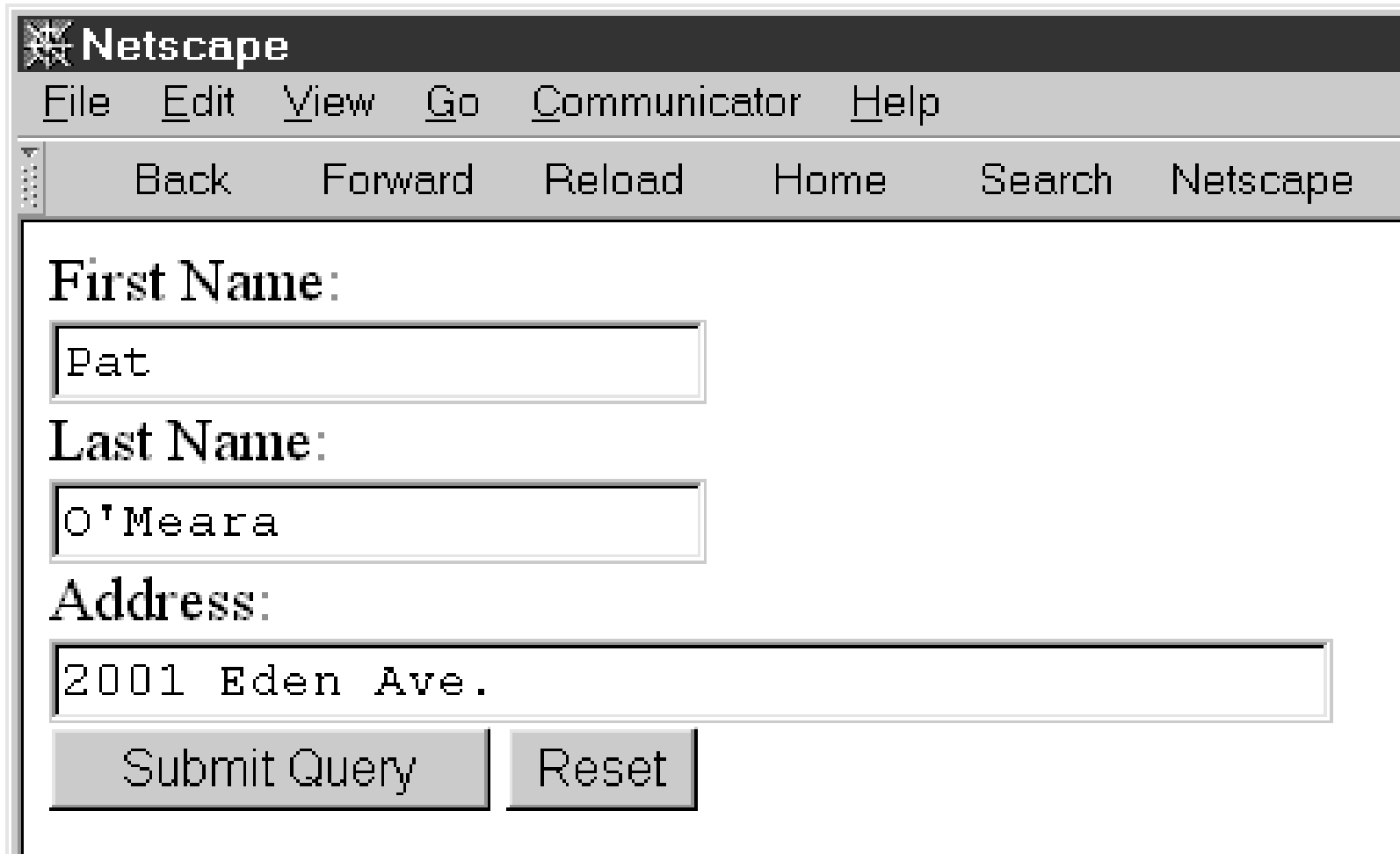
The CGI Process



CGI Scripting - Perl

```
#!/usr/local/bin/perl  
print "Content-Type: text/html\n\n";  
print "<HTML>\n";  
print "Hello World!\n";  
print "</HTML>\n";
```

CGI Form



A screenshot of a Netscape browser window. The title bar reads "Netscape". The menu bar includes "File", "Edit", "View", "Go", "Communicator", and "Help". The toolbar contains "Back", "Forward", "Reload", "Home", "Search", and "Netscape". The main content area displays a form with three text input fields and two buttons. The first field is labeled "First Name:" and contains the text "Pat". The second field is labeled "Last Name:" and contains the text "O'Meara". The third field is labeled "Address:" and contains the text "2001 Eden Ave.". Below the fields are two buttons: "Submit Query" and "Reset".

Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape

First Name:

Pat

Last Name:

O'Meara

Address:

2001 Eden Ave.

Submit Query Reset

CGI Form HTML

```
<HTML>
<FORM ACTION="/cgi-bin/nameform.pl" METHOD="POST">
First Name:<BR><INPUT TYPE=TEXT NAME=fname><BR>
Last Name:<BR><INPUT TYPE=TEXT NAME=lname><BR>
Address:<BR><INPUT TYPE=TEXT NAME=address
SIZE=40><BR>
<INPUT TYPE=SUBMIT> <INPUT TYPE=RESET>
</FORM>
</HTML>
```

Server-Side Includes

- **Not all web sites need CGI**
- **A small amount of data needs to be dynamic**
- **Get the server to fill this in**
- **Use special tags in HTML**
 - **Directives**
 - **Interpreted and replaced with data by the server**

Server-Side Includes

```
<HTML>
```

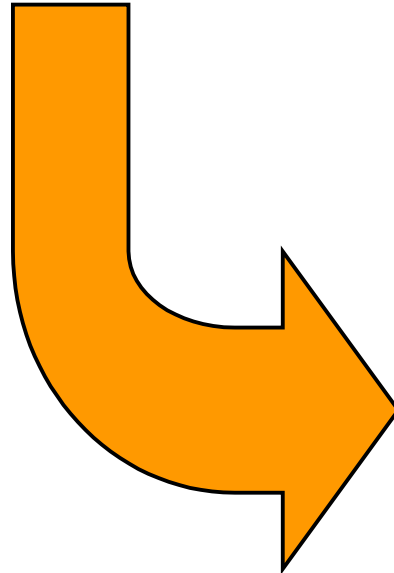
```
My incredibly elaborate web page...
```

```
<HR>
```

```
This page was last updated on:
```

```
<!--#echo var="LAST_MODIFIED"-->
```

```
</HTML>
```



```
<HTML>
```

```
My incredibly elaborate web page
```

```
<HR>
```

```
This page was last updated on:
```

```
06/21/99 10:34:25 PM
```

```
</HTML>
```

Typical Server-Side Includes

● INCLUDE

- Includes a file or a URL at this position
- Useful for maintaining a common look and feel

● EXEC

- Execute a program and insert output at this position

● ECHO

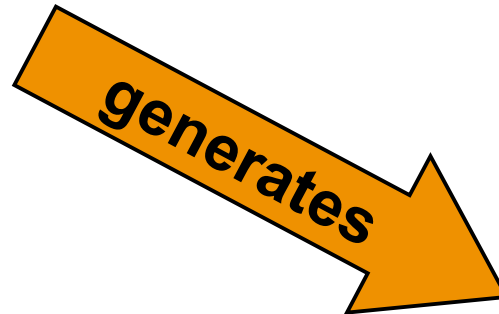
- Insert the value of a variable at this position

Active Server Pages

- **Microsoft proprietary scripting language**
- **Embed scripts into documents**
- **Server process scripts**
 - issues '*pure*' HTML
- **Use any language that supports COM**
 - VBScript, Jscript, C++, Perl, Java

ASP Example

```
<%@ LANGUAGE="VBSCRIPT" %>  
<HTML>  
<% for x = 1 to 6 %>  
<H<% =x %>>Header <% =x %></H<% =x %>>  
<% next %>  
</HTML>
```



```
<HTML>  
<H1>Header 1</H1>  
<H2>Header 2</H2>  
<H3>Header 3</H3>  
<H4>Header 4</H4>  
<H5>Header 5</H5>  
<H6>Header 6</H6>  
</HTML>
```

Servlets and Java Server Pages

- **Sun Microsystems Java Language**
 - a well designed language
 - suitable for both client and server side programming
- **Servlets – Java programs executed by the server**
- **Java Virtual Machine (JVM)**
 - Interprets Java code

Typical Java Code

```
import java.io.*;
import java.util.Date;
import javax.servlet.*;
import javax.servlet.http.*;

public class DateServlet extends HttpServlet {
    public void doGet(HttpServletRequest req,
        HttpServletResponse res)
        throws ServletException, IOException
    {

        res.setContentType("text/plain");
        ServletOutputStream out = res.getOutputStream();
        Date today = new Date();
        out.println(today.toString());

    }
}
```


Java Server Pages

- **Simpler than a full servlet**
- **Similar to SSI**
- **Example:**

```
<HTML>
```

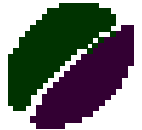
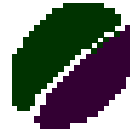
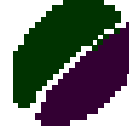
```
Today's date is <%= new Date() %>
```

```
</HTML>
```

- **Code between `<% %>` is executed by server**

Java Beans

- **A component model**
- **Written in Java**
- **Building block for an application**
- **Can be used to provide common functionality wherever required**
- **Similar in principle to Microsoft ActiveX controls**



Summary

● We have covered:

- **Dynamic Documents**
- **CGI and Forms**
- **Server-Side Includes**
- **Active Server Pages**
- **Servlets and Java Server Pages**

Internet Systems Administration

Session 6 Log Files

Contents

- **Log File Formats**
- **Referrers**
- **Being Proactive**
- **Statistics**

Log Files

- **Provide a lot of information about the usage of your web site**
- **You can log any transaction**
- **Debug server-side programs**
- **Help you tune your web site**



Logging Transactions

- **One line per transaction**
- **Not computationally expensive**
- **Can grow very large**
 - **Store on a separate partition**
 - **Rotate log files**
- **Two main logging formats**
 - **Common Logfile Format (CLF)**
 - **Extended Logfile Format (ELF)**

Common Logfile Format

remotehost	rfc1413	authuser	[date]	"request"	status	bytes
------------	---------	----------	--------	-----------	--------	-------

- ***remotehost*** – client hostname or IP address
- ***rfc1413*** – the identity of the remote user (usually -)
- ***authuser*** – users own name (may be -)
- ***[date]*** – date and time of request
- ***request*** – the HTTP request as it came from client
- ***status*** – the HTTP status code returned by the server
- ***bytes*** – the content length of the returned document

Combined Logfile Format

remotehost	rfc1413	authuser	[date]	"request"	status	bytes	referrer	user-agent
------------	---------	----------	--------	-----------	--------	-------	----------	------------

- ***referrer*** – the URL that brought the user to this resource
- ***user-agent*** – the client that made the request
– i.e Netscape 4.5)

Extendible Logfile Format

- Allows the administrator to specify the fields to be logged
- Example:

```
#Version: 1.0
```

```
#Fields: date time c-ip sc-bytes time-taken cs-version
```

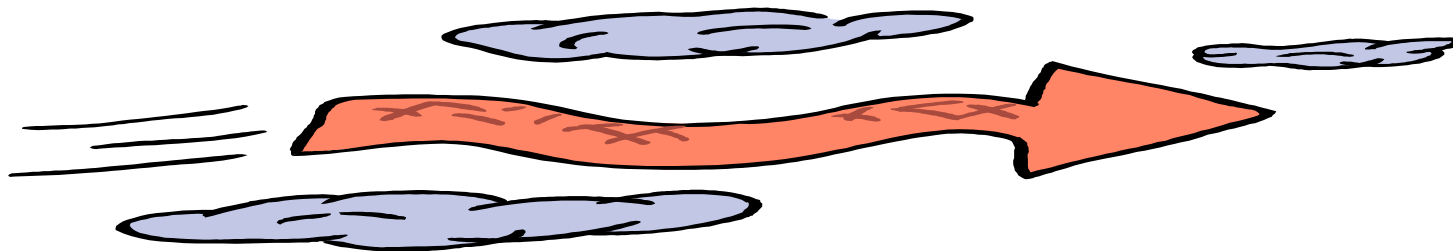
```
1999-08-01 02:10:57 192.0.0.2 6304 3 HTTP/1.0
```

```
1999-08-01 02:12:41 192.0.0.2 5100 1 HTTP/1.0
```

```
1999-08-01 03:37:19 192.0.0.3 5100 2 HTTP/1.0
```

Referrers

- How did visitors reach your web site?
- What web page were they previously viewing?
- Understanding referrers helps you to make your web site more accessible



Referrers

- **Visitor uses a search engine to look for “vortex widgets”**
- **The referrer field might contain:**

Referer:

`http://ink.yahoo.com/bin/query?p=vortex%26widgets&hc=0&hs=0`

Example Error Diagnosis (1)

[Mon Apr 12 19:24:21 1999] [error] Premature end of script headers:

/export/home/patm/public_html/form.cgi

syntax error at form.cgi line 7, near ") print"

Execution of form.cgi aborted due to compilation errors.

Error is in line 7

As a result the form did not work

Example Error Diagnosis (2)

**[Sun Apr 18 16:40:40 1999] [error] Permission denied:
file permissions deny server access:
/export/home/patm/public_html/phonelist.txt**

**Web server
is denied
access**

**[Mon Apr 12 19:43:45 1999] [crit] Permission denied:
/opt/apache/share/htdocs/wml05/class6/.htaccess
pcfg_openfile: unable to check htaccess file, ensure
it is readable**

**[Mon Aug 9 21:55:53 1999] [error] [client
24.218.82.54] access to /sales/ failed, reason: user
erici not allowed access**

**User
denied
access**

Statistics

- **Getting information from data**
- **Understand usage patterns for your web site**
- **Obtain “hit counts”**
- **Tune your web site to attract more users**

Log File Analysis

- **Most requested pages**
- **Top entry pages**
- **Most used browsers**
- **Bandwidth utilisation**
- **Most active domains**
- **Information about search engines**
- **Top referring sites and URLs**
- **Error counts**



Statistics

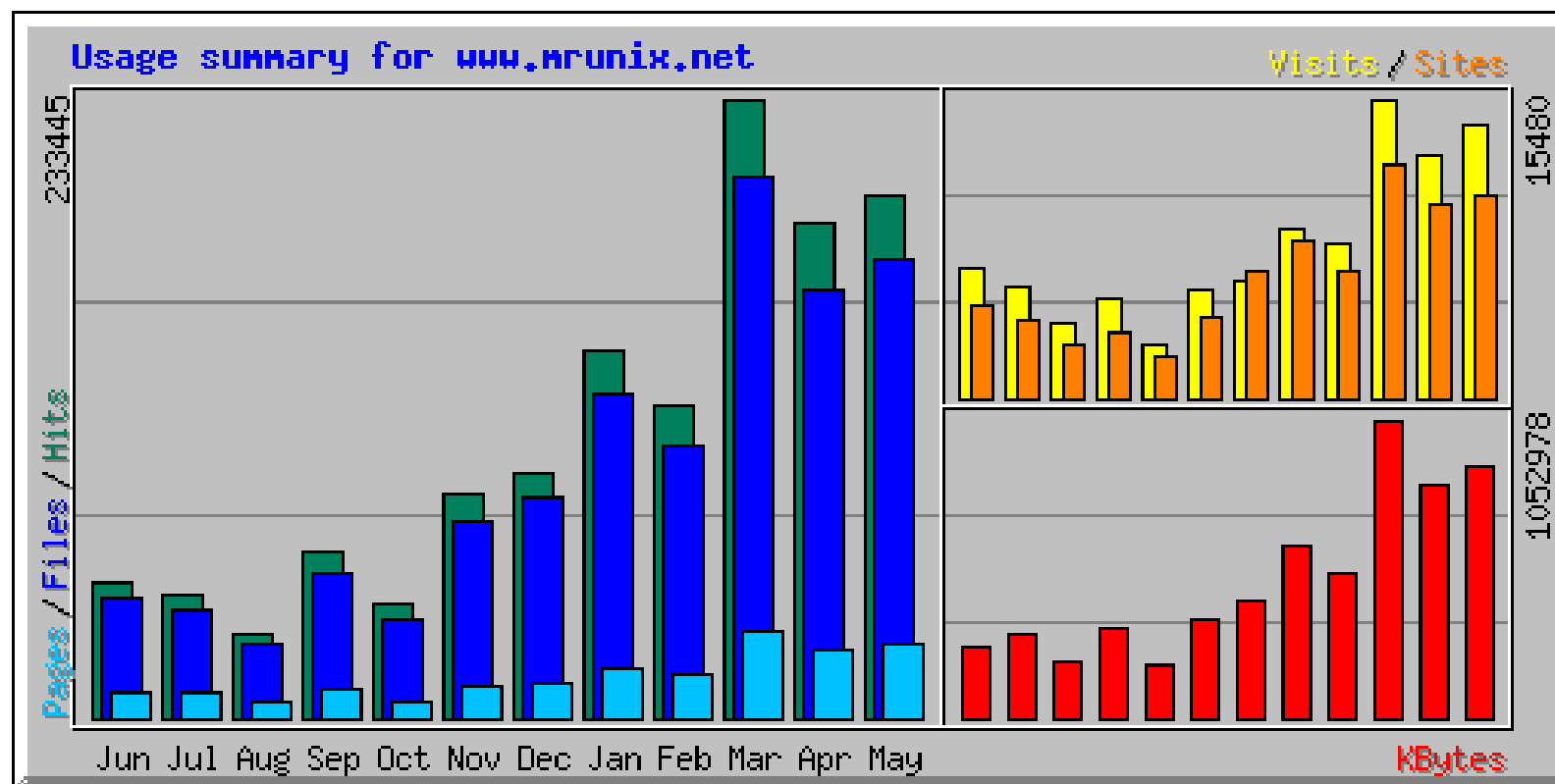
- You need tools to extract statistics from logs
- Webalizer
 - <http://www.mrunix.net/webalizer>
- WebTrends
 - <http://www.webtrends.com>
- Wusage
 - <http://www.boutell.com/wusage>

Using Webalizer (1)

Usage Statistics for www.mrunix.net

Summary Period: Last 12 Months

Generated 23-Oct-2000 02:17 EDT



Using Webalizer (2)



Summary by Month										
Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
<u>May 1999</u>	6377	5570	903	455	10484	884568	14119	28004	172671	197696
<u>Apr 1999</u>	6216	5394	858	419	10087	821968	12594	25758	161844	186504
Mar 1999	7530	6582	1046	499	12128	1052978	15480	32432	204059	233445
Feb 1999	4712	4128	656	321	6629	511793	8048	16419	103203	117816
Jan 1999	4470	3934	607	284	8079	605694	8808	18844	121980	138571
Dec 1998	2998	2673	411	197	6524	410110	6120	12769	82875	92951
Nov 1998	2910	2567	400	192	4260	346705	5588	11627	74468	84403
Oct 1998	3052	2668	457	202	2203	189253	2839	6399	37360	42738
Sep 1998	2072	1826	345	169	3475	314492	5075	10376	54807	62165
Aug 1998	1014	901	211	125	2693	196560	3890	6571	27958	31455
Jul 1998	1484	1325	302	184	4041	298225	5716	9383	41102	46019
Jun 1998	1707	1502	322	222	4809	251502	6675	9687	45077	51227
Totals						5883848	94952	188269	1127404	1284990

Summary

● We have covered:

- Log File Formats
- Referrers
- Being Proactive
- Statistics

Internet Systems Administration

Session 7

Search Engines, Robots and Automation

Contents

- **Search Engines**
- **Publicising your site**
- **Robots and Spiders**
- **Automation**

Objectives

- **How to increase usage of your site**
- **How to enhance it**
- **How to publicise it**
- **How your site interacts with search engines**

Search Engines

- **How can users find content within your site?**
- **Navigation bar or table of contents**
 - simple but slow
- **A web site based search engine**
 - greatly increases usefulness of your web site
 - easy to install
 - offers other useful features

Search Engines

- **Excite for Web Servers (EWS)**
 - <http://www.excite.com/navigate/>
- **SWISH-E (Simple Web Indexing System for Humans –Enhanced)**
 - <http://sunsite.berkeley.edu/SWISH-E/>
- **AltaVista search Intranet**
 - <http://www.altavista-software.com/>
- **Microsoft Index Server**
 - <http://www.microsoft.com/NTServer/fileprint/exec/feature/Indexfaq.asp>

Example – SWISH

 Berkeley Digital Library SunSITE..... Home..... Search..... Index

Search the SunSITE

Search:

A Boolean "and" between words is assumed; that is, documents will be retrieved that have **all** the specified words. If you wish to find all the documents that have **any** word, then use "or" between your search words.

To truncate a word, use an asterisk (*) at the end. For example, the search "comput*" would retrieve documents that have the words "computer", "computing", "computation", etc.

Powered by

SWISH-E

.....

[Copyright © 1998 UC Regents](#). All rights reserved.

Document created at the moment of request by <http://sunsite.berkeley.edu/cgi-bin/search.pl>

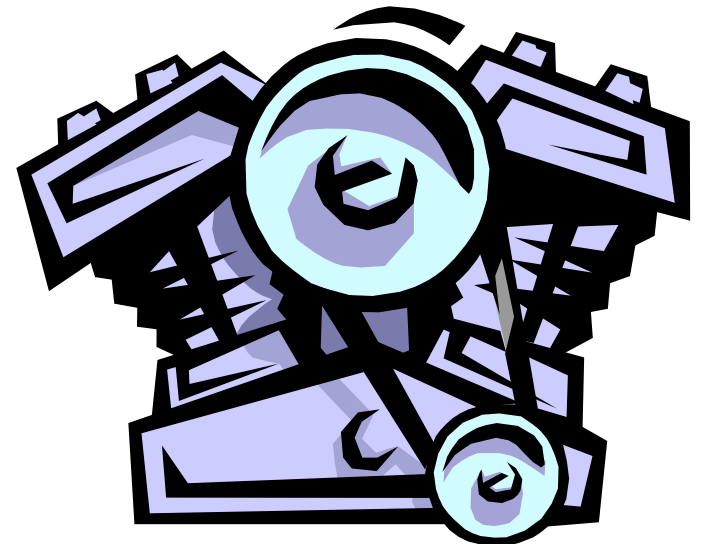
Service maintained by the SunSITE Manager: manager@sunsite.berkeley.edu

Virtual Search Engine

- Use an existing high capacity search engine
 - Infoseek
- Filter out all hits *except* those for your web site
- No cost, no disk space, minimal effort
- BUT
 - Only if your site is properly indexed by the search engine

Installing a Search Engine

- **Indexing your site**
- **Search Form**
- **Search Engine**



Publicising Your Site

- Use tags well

- <META>

- Use keywords to describe your site and capture hits

- <TITLE>, <H1>, <H2>

- Search engines look at these



Publicity Tips

- **Register with the major search engines and directories**
 - Search engines match key words
 - Directories match topics
 - Metasearch engines use other search engines
- **Content needs to be useful/interesting**
- **Advertise your site**
- **Find sites that will link to yours**

Controlling the Robots and Spiders

- By default robots visit every page of your site
- This may be what you want, but:
 - what about dynamic content?
 - what if you have too many matching keywords?
- Use a robot exclusion protocol



Robots.txt

My robots.txt file

User-agent: *

Specify which
robots are
controlled

Disallow: /cgi-bin

Disallow: /development

Disallow: /beta

Specify which
directories are
not allowed

Automation

- **System administrators have a lot of responsibility**
 - managing disk space
 - checking for errors in the logs
 - generating reports
 - performing backups
- **Scripts can automate most of these**



UNIX Tools (1)

- **UNIX has strong tools and scripting languages**
- **CRON**
 - Daemon that starts programs according to a schedule
 - Crontab creates list of tasks
- **AT**
 - Run a command at a future time
 - Windows NT has a version of AT



UNIX Tools (2)

- **Perl, Tcl**
 - very powerful, easy to use scripting languages
- **Shell scripts**
 - csh, ksh, etc
 - files containing OS commands and programming features
- **Expect**
 - automation of command-line tasks where there is user interaction in response to system prompts



Summary

- **We have covered:**
 - **Search Engines**
 - **Publicising your site**
 - **Robots and Spiders**
 - **Automation**